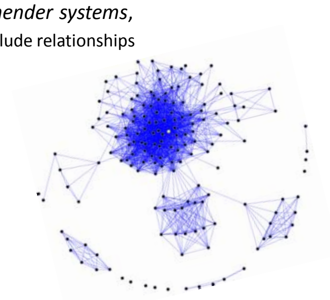
CS6120:
Intelligent Media Systems

Dr. Derek Bridge
School of Computer Science & Information Technology
UCC

Social Recommender Systems

- In *social recommender systems*,
  - user profiles include relationships between users



MODIFICATIONS TO COLLABORATIVE RECOMMENDERS

k-Nearest Neighbours for User-Based Collaborative Recommending

for each candidate item, i
  for each user v in $\mathbb{U}_i$ except for active user u, i.e. other users who have rated i
    compute $sim(u, v)$
  let NN be the set of k nearest neighbours, i.e. the k users who have rated i and who are most similar to u and whose similarity to u is positive
  let $\widehat{r_{u,i}}$ be the *weighted average* of the ratings for i in NN
recommend the candidates in descending order of predicted rating

Modification: Friends

for each candidate item, i
  for each user v **who is a friend of u** and has rated i
    compute $sim(u, v)$
  let NN be the set of k nearest neighbours, i.e. the k users who have rated i and who are most similar to u and whose similarity to u is positive
  let $\widehat{r_{u,i}}$ be the *weighted average* of the ratings for i in NN
recommend the candidates in descending order of predicted rating

Modification: Friends

Problem:
too few friends
- Coverage and accuracy problems
  - not enough friends to make good predictions

Problem:
same bed, different dreams
- The interest graph is not equal to the social graph

Modification: Trust

for each candidate item, i

for each user v in $\mathbb{U}_i$ except for active user u, i.e. other users who have rated i

compute $\boldsymbol{trust(u, v)}$

let NN be the set of k nearest neighbours, i.e. the k users who have rated i and who u trusts most and whose trust by u is positive

let $\widehat{r_{u,i}}$ be the *weighted average* of the ratings for i in NN

recommend the candidates in descending order of predicted rating

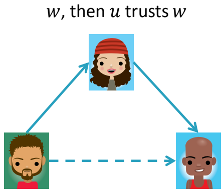
Explicit Trust vs Implicit Trust

- Explicit
  - In some systems, users select who they trust
    - e.g. Epinions
  - NB $trust(u, v)$ means u trusts v's opinion
- Implicit
  - In other systems, trust is inferred from user interaction
    - likes, shares, comments, visits to profile, chats,...

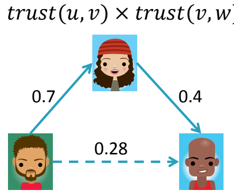
- But your immediate web of trust is small
  - coverage and accuracy problems

Trust Propagation

- Extend your web of trust using transitivity
  - if u trusts v and v trusts w, then u trusts w

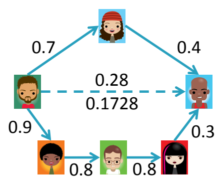


- Quantify this, e.g. by multiplication
  $trust(u, w) = trust(u, v) \times trust(v, w)$



Trust Aggregation

- Aggregation
  - for quantifying the trust when there is more than one path from u to w
  - Options
    - use the score of the shortest path
    - use the highest score
    - average the scores
    - …



Discussion

Using user-user similarity

1. Sparsity can reduce coverage and accuracy
2. Cold-start problem with new users who have few ratings
3. Prone to attacks
4. The anonymity makes this less transparent

Using trust

1. Trust reaches more users (if it is propagated enough)
2. Reduced cold-start problem:
   - each new trust relationship brings more coverage and accuracy than each new rating
   - can import your social graph from another site
3. Attacks are more difficult
4. More transparent: more like word-of-mouth

Discussion

- **Trust**
  - a relationship between two users
- **Reputation**
  - the community's view of a user
  - e.g. by aggregating trust or by votes on reviews/ comments/ answers to questions
  - e.g. karma in reddit
  - e.g. reputation and badges in stackoverflow
  - e.g. an academic's H-score
- In a user-based collaborative recommender you could use a combination of
  - user-user similarity
  - trust
  - reputation
- In an item-based collaborative recommender, you could weight the users by trust or reputation within the item-item similarity

NEWSFEED PRIORITIZATION

Facebook NewsFeed

- Stories from your friends
  - the average user has 1500 candidate stories
  - so Facebook must show you a subset
  - it scores them and shows those with highest scores
- Facebook terminology
  - Object: a status update, a comment, a fan page,…
  - Edge: an interaction with an object
    - creating it, liking it, commenting on it, sharing it, hiding it, stopping notifications, friending a person, tagging a photo, adding a location to a photo, joining an event,…

EdgeRank

- An object is more likely to show up in your NewsFeed if people you know have been interacting with it recently
- Score will be higher
  - the better you know them
  - the more they interact with it
  - the more recent the interaction

From edgerank.net

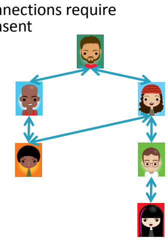
NewsFeed Prioritization

- EdgeRank
  - Affinity, u_e:
    - a measure of your engagement with this friend or page
  - Weight, w_e:
    - e.g. commenting has more weight than liking
    - e.g. changes to relationship status have high weight
  - Time decay factor, d_e:
    - recency of activity
- Since 2010, Facebook uses 100,000 signals
  - family vs close friend vs acquaintances vs friends of friends
  - number of friends interacting with an object; total number of people interacting with an object
  - type of post that you seem to interact with more
  - device; speed of Internet connection
  - …

WHO TO FRIEND OR FOLLOW

Facebook: People You May Know

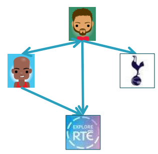
- Social graph
  - friends and family
  - edges in the graph are bidirectional because connections require consent



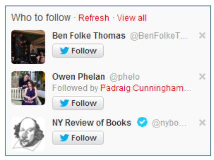
- Recommendations
  - ideally people you might know, not total strangers
  - primarily based on number of mutual friends that you interact with a lot
- From Facebook:
  - "We show you people based on mutual friends, work and education information, networks you're part of, contacts you've imported and many other factors."

Twitter: Who To Follow

- Interest graph
  - more connections to sources of info than to friends & family
  - edges in the graph are often unidirectional



- Recommendations
  - can be 'strangers'
  - users you might be similar to
  - users you might be interested in



General Link Prediction Methods

- These are methods based on the structure of the graph
  - e.g. nodes represent users
  - e.g. edges represent relationships (friend, follower) or interactions
  - obviously cannot take into account extrinsic factors, e.g. a user relocates and becomes a (physical) neighbour of another
- If there is no edge between nodes u and v, these methods give a score
  - the missing edges with the highest scores are recommended

General Link Prediction Methods

Shortest paths

- Find the shortest path in the graph from u to v
- Calculate its length, l
- The score of the missing edge from u to v is negative l

Neighbourhoods

- Let $nbrs(u)$ and $nbrs(v)$ be the neighbours in the graph of u and v
- The score of the missing edge from u to v is, e.g.
  - size of intersection: $|nbrs(u) \cap nbrs(v)|$
  - Jaccard: $\frac{|nbrs(u) \cap nbrs(v)|}{|nbrs(u) \cup nbrs(v)|}$
  - preferential attachment: $|nbrs(u)| \times |nbrs(v)|$

General Link Prediction Methods

All paths

- Find all paths in the graph from u to v
- Let p_l be the set of paths from u to v that are of length l
- Katz: the score of the missing edge from u to v is $\sum_{l=1}^{\infty} \beta^l \times |p_l|$ where $0 < \beta \leq 1$

Random walks

- Inspired by PageRank
- Randomly surf
  - starting at, e.g., u
  - count how many times you visit the other nodes such as v
  - use this as the score for the missing edge

WTF: Twitter's Early Method

- First, find the active user's circle-of-trust
  - found by a kind of personalized PageRank
    - start at the active user
    - randomly walk the graph
    - teleportation takes you back to the active user
    - score users by how often they get visited
- Let *C* be the users in the circle-of-trust
- Let *F* be the users who are followed by users in *C*
- Next, use a random walk algorithm called SALSA
  - start at a user in *C*
  - repeatedly
    - traverse a link to a user in *F*
    - traverse a link back to a user in *C*
- Last, recommend users in *C* and *F* with high scores
  - users in *C* are similar to the active user
  - users in *F* are interesting to the active user

Twitter's Current Method

- There are no real details!
- It is a weighted hybrid (ensemble) of 20+ scoring methods
  - dynamically adjusts the weighting
- As well as graph structure, they also use:
  - data about nodes (users), e.g. how many times a user has retweeted
  - data about the edges, e.g. timestamps
  - other kinds of edges (interactions), e.g. retweets, favourites, replied
- They could use, but probably don't:
  - tweet content
- They certainly do:
  - lots of A/B testing!

Ongoing Research

- Similarly, recent research papers have discussed data that is extrinsic to the graph
  - Facebook: demographics
  - LinkedIn: overlap in the time two people belonged to an organization, organization size, role,…
- Unclear how much of this is in their current algorithms

GROUP RECOMMENDER SYSTEMS

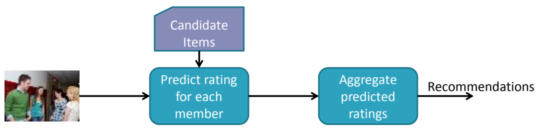
Consuming Items Together





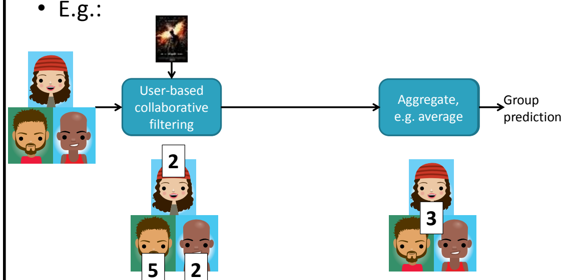

Aggregation of each Group Member's Predicted Rating

- Group recommenders:
  - the most common approach:



Aggregation of each Group Member's Predicted Rating

- E.g.:



Rating Aggregation Methods

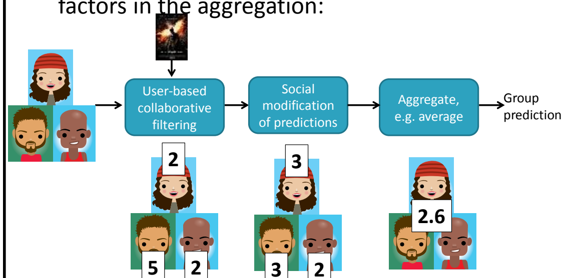
- Average:
  - the mean predicted rating for the item
- Median:
  - the middle predicted rating for the item
- Multiplicative:
  - multiply the predicted ratings
- Least Misery:
  - the minimum predicted rating
- Most Pleasure:
  - the maximum predicted rating
- Borda count:
  - sum the inverse ranks

Issues in Aggregation

- Normalize each user's ratings before prediction and aggregation
- Manipulation of outcome
  - if ratings are explicit ones, users can see others' ratings, and they know the aggregation method
  - how would you manipulate Least Misery?
  - what about Median?
  - on the other hand, in this scenario, you sometimes get conformity effects
- People expect fairness over time
  - if the same group gets a subsequent recommendation, a member who 'lost out' first time should carry more weight
  - but group membership often varies a bit over time
- Social factors influence the item the group will agree on
  - personalities
  - relationships within the group

Issues in Aggregation

- A group recommender could anticipate social factors in the aggregation:



Issues in Group Recommenders

- Explanations
- Active versus passive groups
  - Active:
    - e.g. negotiating over a movie/ vacation
  - Passive:
    - e.g. choosing music or temperature for a shared space such as a shop, gym or office
- What does a rating mean?
  - my opinion of the item
  - my opinion of the item for this group
  - my opinion of the experience